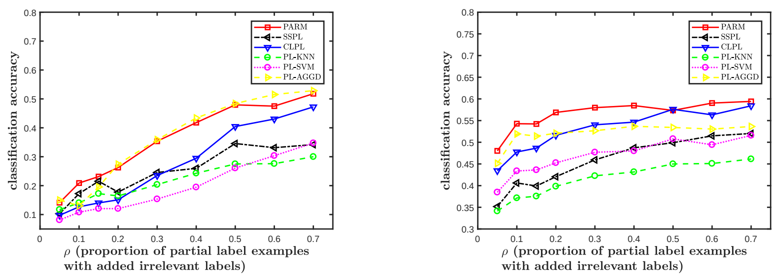
First of all, we wish to sincerely thank the anonymous reviewers for their time and efforts in reviewing our NeurIPS submission #5474. Next, we would like to provide responses to major concerns raised in the reviewing comments:

[Limited novelty]

In this paper, the first maximum margin solution towards the problem of semi-supervised partial label learning is proposed. To the best of our knowledge, the SSPL [22] approach corresponds to the only prior work on the same problem studied in this paper. The key differences between SSPL and the proposed PARM approach correspond to: 1) SSPL employs graph-based label propagation for estimating the labeling confidence over both partial label and unlabeled examples, while PARM employs label propagation to instantiate the labeling confidences over partial label examples. The labeling confidences over unlabeled examples are estimated by PARM based on follow-up maximum margin procedure; 2) Due to the transductive nature of graph-based methods, SSPL is not meant to be able to make predictions on unseen examples during testing phase. As a remedy, SSPL further applies kNN rule over training examples with estimated labels to enable inductive prediction on unseen examples. Due to the inductive nature of maximum margin approach, PARM is capable of making predictions on unseen examples without resorting to extra procedure. In the revised version, we will make this clearer in the "Related Work" section.

[Variable sizes of candidate label set]

To illustrate the performance of PARM on datasets with larger and variable size of candidate label set, we enlarge the candidate label set of partial label examples in `Lost` and `BirdSong` datasets by randomly adding irrelevant labels into their candidate label set. Consequently, by increasing the proportion (ρ) of partial label examples with randomly added irrelevant labels, the size of candidate label set would vary from 8 to 10 for `Lost` dataset and from 5 to 9 for `BirdSong` dataset respectively. Figure 1 illustrates how PARM and the comparing approaches perform as ρ increases from 0.05 to 0.7. The results clearly show the advantage of PARM in learning from partial label examples with larger and variable size of candidate label set.

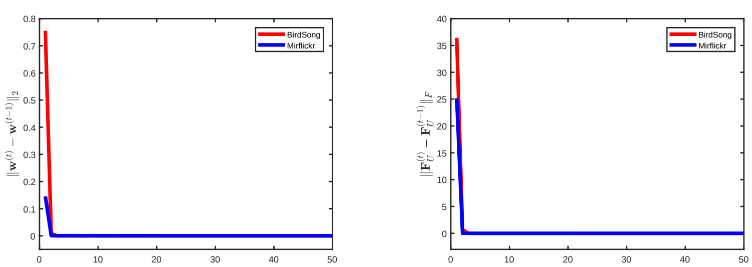


(a) On `Lost` data set (b) On `BirdSong` data set

Figure 1: Classification accuracy of PARM and each comparing approach with varying size of candidate label set.

[Convergence analysis]

Figure 2 illustrates how the classification model (i.e. $\|\boldsymbol{w}^{(t)} - \boldsymbol{w}^{(t-1)}\|_2$) and the confidence matrix over unlabeled examples (i.e. $\|\mathbf{F}_U^{(t)} - \mathbf{F}_U^{(t-1)}\|_{\mathrm{F}}$) converge as the number of optimization iterations t increases. The high convergence rate of PARM is desirable for dealing with data sets with larger scale.



(a) $\boldsymbol{w}$ (b) $\mathbf{F}_U$

Figure 2: Convergence curves of $\boldsymbol{w}$ and $\mathbf{F}_U$ (on `BirdSong` and `Mirflickr`).

[Definition of σ]

The parameter σ corresponds to the width of Gaussian kernel, which is fixed to be 1 in this paper (pp.3, footnote 1).